

DOT/FAA/AM-02/24

Office of Aerospace Medicine
Washington, DC 20591

The Effects of Practice and Coaching on the Air Traffic Selection and Training Test Battery

Michael C. Heil
Cristy A. Detwiler
Rebecca Agen
Clara A. Williams
Brandy O. Agnew
Raymond E. King
Civil Aerospace Medical Institute
Federal Aviation Administration
Oklahoma City, OK 73125

December 2002

Final Report

DISTRIBUTION STATEMENT A
Approved for Public Release
Distribution Unlimited

This document is available to the public
through the National Technical Information
Service, Springfield, VA 22161.



U.S. Department
of Transportation
**Federal Aviation
Administration**

20030123 050

N O T I C E

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents thereof.

Technical Report Documentation Page

1. Report No. DOT/FAA/AM-02/24	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle The Effects of Practice and Coaching on the Air Traffic Selection and Training Test Battery		5. Report Date December 2002	
		6. Performing Organization Code	
7. Author(s) Heil MC, Detwiler CA, Agen RA, Williams CA, Agnew BO, and King RE		8. Performing Organization Report No.	
9. Performing Organization Name and Address FAA Civil Aerospace Medical Institute P.O. Box 25082 Oklahoma City, OK 73125		10. Work Unit No. (TRAIS)	
		11. Contract or Grant No.	
12. Sponsoring Agency name and Address Office of Aerospace Medicine Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591		13. Type of Report and Period Covered	
		14. Sponsoring Agency Code	
15. Supplemental Notes This report was performed under Task HRR521.			
16. Abstract The Air Traffic Selection and Training (AT-SAT) test battery is the Federal Aviation Administration's (FAA's) recently developed computerized selection test for Air Traffic Control Specialists (ATCSs). Only one form of the AT-SAT battery was developed as part of the initial development and validation effort, meaning that all people who take the test receive the exact same items. The goals of the current study were to: 1) determine if repeated test taking improves performance; 2) determine if coaching improves performance; 3) identify specific tests within the AT-SAT battery that are most susceptible to practice and coaching effects; and, 4) determine the extent to which practice and coaching effects potentially impact hiring decisions. Study participants were not ATCSs; however, they had to meet basic requirements for the ATC occupation to be eligible for participation. They were recruited through a contractor and randomly assigned to one of three experimental groups. Group 1 received a one-day coaching intervention before taking the first administration of AT-SAT. Group 2 took the first administration of AT-SAT, and then received the coaching intervention before the second administration. Group 3, the control group, took AT-SAT three times without coaching. Test scores were compared both between and within each group using ANOVA with repeated measures. The results suggest that performance on the AT-SAT battery may indeed be influenced by both practice and coaching effects. More specifically, the results demonstrate that the composite AT-SAT score that is used for hiring decisions increases with repeated administrations, although the greatest increase occurs following coaching. In terms of selection decisions, it is conceivable that coaching could move an individual from a failing status into a passing status and even from the qualified category into the well-qualified category, without improving their ability to perform on the job.			
17. Key Words AT-SAT, Air Traffic Control, Coaching Effects, Practice Effects, Personnel Selection		18. Distribution Statement Document is available to the public through the National Technical Information Service; Springfield, VA 22161	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 13	22. Price

THE EFFECTS OF PRACTICE AND COACHING ON THE AIR TRAFFIC SELECTION AND TRAINING TEST BATTERY

The Air Traffic Selection and Training (AT-SAT) test battery is the Federal Aviation Administration's (FAA's) recently developed computerized selection test for Air Traffic Control Specialists (ATCSs). The AT-SAT project was initiated in October 1996 to address the FAA's need for a new selection instrument. The purpose of the project was to develop a valid, legally defensible, job-related, computerized ATCS selection battery. The new selection test battery is intended to screen ATCS applicants; those who are selected based on their score will be hired by the FAA and sent to the Academy for training. The AT-SAT test battery is based on the Separation and Control Hiring Assessment (SACHA) job analysis (Nickels, Bobko, Blaine, Sand & Tartak, 1995).

Only one form of the AT-SAT battery was developed as part of the initial development and validation effort, meaning that all people who take the test receive the exact same items. Consequently, there is an increased likelihood that any improvement in the score of someone who retakes the test is due to a practice effect. The use of one form also suggests that the test may be more vulnerable to coaching since there is only one set of items that must be trained. The goals of the current study were to: 1) determine if repeated test taking improves performance; 2) determine if coaching improves performance; 3) identify specific tests within the AT-SAT battery that are most susceptible to practice and coaching effects; and, 4) determine the extent to which practice and coaching effects potentially impact hiring decisions.

The distinction between practice and coaching has been made in previous research (Mauer, Solamon, & Troxtel, 1998). Practice can be defined as taking a particular test multiple times so that the test taker becomes familiarized with the format of the test; while coaching can be defined as an intervention that provides suggestions for the test taker to improve his/her test performance. Additionally, five types of coaching interventions have been identified (Sackett, Burris, & Ryan, 1989). They include:

- an intensive drill on items similar to that of the selection device;
- giving specific tips on test taking, which are not related to the content of the test;
- orienting the test taker to the selection device itself;

- giving the test taker principles for dealing with the content of the test; and,
- giving behavioral advice for dealing with the selection device.

The distinction between practice and coaching lies in whether or not there is an outside intervention.

Research has shown that both practice and coaching have significant effects on many of the widely used and marketed standardized tests on the market today (Powers, 1993). For example, the Educational Testing Service concedes that practice alone can increase the Scholastic Aptitude Test (SAT) verbal score by 15 points and the math score by 12 points, while coaching can increase both verbal and math scores by 15-25 points. Although extensive research has been done on how practice and coaching can affect standardized test scores, the question that remains is how might practice and coaching influence the employment selection process. When the accuracy of a selection device is compromised, less informed hiring decisions are inevitably made. Specifically, if practice and/or coaching increases an applicant's score on a selection test, which leads to inferences about the amount of a characteristic being measured, then the predictive validity of that selection device is undermined (Sackett et al., 1989).

Research has indicated that the employment selection process is not immune to the effects of practice and coaching. Mauer et al. (1998) illustrated that coaching is related to successful performance in structured interviews, which are a widely used method of selection. Alliger, Lillienfeld, and Mitchell (1996) reiterated that coaching on overt integrity tests is related to significant increases in test scores. Moreover, tests that measure personality or temperament, also used in some organizations for the purpose of selection, have been made available to the public so that people can practice (Furhnam, 1997).

The implications for personnel selection include concerns about how practice and coaching might interfere with the ability of a test or test battery to accurately predict future job performance. In addition, there is a concern that test scores that improve due to practice and coaching effects might influence hiring decisions based on test performance.

A study of coaching and practice effects using an earlier version of the Air Traffic Scenarios test, which is included in the AT-SAT battery, demonstrated that test strategy training may improve performance on the test (Gilliland & Schlegel, 1992). Analysis of archival data revealed that ATCSs who had taken the Office of Personnel Management (OPM) selection test multiple times improved their OPM score yet did not perform as well as other ATCSs on AT-SAT concurrent validation criterion measures such as the Computer-Based Performance Measure (Manning & Heil, 2001). This suggests that improvement of a test score due to repeatedly taking the test did not result in a change in underlying cognitive skills.

One hypothesis of the current study was that participants' overall score on the AT-SAT battery would improve with repeated trials and that there would be a greater increase in overall scores for participants who were taught specific strategies for enhancing performance. It was further hypothesized that tests that were based on computer simulation performance—the Air Traffic Scenarios, Scan, and Letter Factory tests—would be more vulnerable to practice effects than other tests contained in the battery. This hypothesis is based partly on the results of the Gilliland and Schlegel (1992) study cited above. Another basis for the hypothesis is that people who are repeatedly exposed to the simulation tests have the opportunity to observe patterns and practice strategies that may help them improve their performance during later test administrations.

METHOD

Participants

Study participants were recruited through a contractor and were subsequently paid for their participation. To be eligible for participation, participants had

to meet several requirements: (a) be U.S. Citizens between the ages of 18 and 30, (b) have normal color vision, (c) have a high school diploma or equivalent, (d) have the ability to operate a computer mouse, and, (e) have no prior air traffic control or direct aviation experience (e.g., pilot, airline dispatcher, crew chief, bombardier, or navigator). Points a, b, and c above are requirements for the ATCS job. A total of 150 participants were recruited. Participants were randomly assigned to one of three experimental groups. As shown in Table 1, each group had a relatively equal proportion of males and females, with males making up slightly more than half of each group. The educational level of participants is summarized in Table 2.

Measures

The AT-SAT test battery is a newly developed, computerized test of cognitive ability. The AT-SAT battery is comprised of seven tests of cognitive ability and one non-cognitive measure. In addition to the AT-SAT composite score based on total test performance, scores were also calculated for each sub-test included in the battery. A description of each sub-test is given in Appendix A.

Procedure

During a pre-screening session conducted prior to participating in the study, participants filled out a background questionnaire, the NEO Personality Inventory, and an informed consent form. All three groups took AT-SAT a total of three times with an interval of three weeks between each testing session. Group 1 received a one-day coaching intervention before taking the first administration of AT-SAT. Group 2 took the first administration of AT-SAT, and then received the coaching intervention before the second administration. Group 3, the control group,

Table 1. Participants in Each Experimental Condition

Experimental Condition	N	Male	Female
Coaching prior to testing	47	55.3% (26)	44.7% (21)
Coaching after testing	47	57.4% (27)	42.6% (20)
Control	49	57.1% (28)	42.9% (21)

Table 2. Educational Level of Participants

Experimental Condition	High School	Trade School	Attended College	College Degree	Graduate School
Coaching prior to testing	14.9% (7)	6.4% (3)	63.8% (30)	12.8% (6)	2.1% (1)
Coaching after testing	36.2% (17)	4.3% (2)	48.9% (23)	10.6% (5)	0% (0)
Control	14.6% (7)	12.5% (6)	54.2% (26)	10.4% (5)	8.3% (4)

took AT-SAT three times without coaching. This design, presented in Table 3, allowed for the measurement of practice (repeated testing) effects, coaching effects, as well as the practice/coaching interaction.

Participants were tested on ten computer workstations, separated by partitions, which allowed ten participants to be tested simultaneously. Participants received identical instructions before each test administration, and had a schedule that allowed for two 15-minute breaks and a 45-minute lunch break.

The one-day (6-7 hour) coaching intervention was conducted in a classroom environment and given the day before a test administration. Developed by Air Traffic Control instructors and Civil Aerospace Medical Institute (CAMI) researchers, the coaching curriculum was presented as a PowerPoint presentation and provided an overview of each subtest, sample items, strategies for taking that particular subtest, and overall test-taking strategies. Handouts were not given to participants, and they were not allowed to take notes. Coaching always took place on the day prior to test administration. Each group of participants took the AT-SAT battery a total of three times, with three weeks between each session.

RESULTS

Group Means

Test scores were compared both between and within each group using ANOVA with repeated measures. The results, summarized in Table 4, show that the mean weighted composite AT-SAT score was significantly higher for participants who had received coaching ($F = 55.00$ (2, 115), $p < .01$). The main effect for

within subjects comparisons (practice) was also significant ($F = 55.00$ (2, 115), $p < .01$). After the first administration of AT-SAT, participants who received coaching prior to taking the first test scored significantly higher than participants from the two groups that had not yet received coaching. Following the second test administration, the scores for people who received coaching after they had taken the test once increased and were significantly higher than the Time 2 scores of the control group, people who had not received coaching. Although the scores for all groups increased after the second test administration, the increase was greater for those who had just received coaching. Although the mean Time 2 AT-SAT score for people who were coached prior to the second test administration remained lower than that of people who had been coached prior to Time 1, the difference was not statistically significant. The results of the ANOVA suggest that the composite test score, as well as scores on several of the tests that comprise the battery, were influenced by both coaching and practice. The mean composite AT-SAT scores for each group at each time of testing are presented in Table 4 and plotted in Figure 1. The results of the ANOVA for each AT-SAT sub-test are presented in Table 5.

Practice and Coaching Index

Direct comparisons of all AT-SAT sub-tests were performed by calculating a practice index and a coaching index for the AT-SAT composite and each score. Since the AT-SAT sub-tests are all scored on different scales, it was necessary to convert the scores generated by these tests to a standard score so that these comparisons could be made. Both the practice and the coach-

Table 3. Testing Schedule

Group	Treatment
Coaching Prior to testing	Coach → Test 1 → Test 2 → Test 3
Coaching After Testing	Test 1 → Coach → Test 2 → Test 3
Control	Test 1 → Test 2 → Test 3

Notes. Coaching always occurred the day before the next testing session.
All test sessions were 3 weeks apart.

Table 4. Mean AT-SAT Score for Each Experimental Group

Group	1	2	3
Control	68.0*	75.3*	76.6*
Coaching After Testing	71.0*	80.2	80.6
Coaching Prior to Testing	82.2	84.7	86.0

* Mean scores that are significantly different than the *coaching prior to testing* group mean ($p < .05$).

Figure 1. Plot of AT-SAT Weighted Composite Score

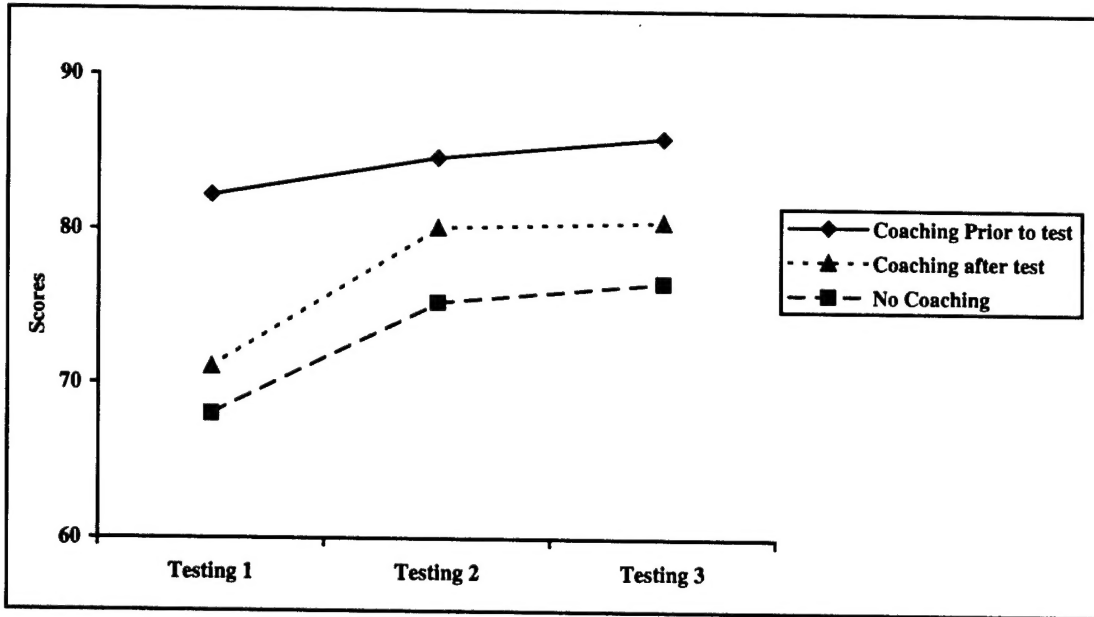


Table 5. Mean Test Scores for Each Experimental Group

Test	Session 1			Session 2			Session 3			F
	No Coaching	Coaching after Testing	Coaching Prior to Testing	No Coaching	Coaching after Testing	Coaching Prior to Testing	No Coaching	Coaching after Testing	Coaching Prior to Testing	
Applied Math	8.0 ^b	8.3	9.9 ^a	8.6	8.6	10.3	8.7 ^b	8.8 ^b	10.9 ^a	8.29
Angles	13.1 ^b	12.9 ^b	15.3 ^a	13.9 ^b	14.1	15.4 ^a	14.0 ^b	14.5	15.6 ^a	12.48
LF-Sit. Aware.	2.3 ^b	2.4 ^b	2.9 ^a	2.8 ^b	3.1	3.4 ^a	2.9	3.3	3.4	49.25
LF-Plan Ahead	4.8 ^b	4.8 ^b	5.6 ^a	4.6 ^b	5.3	5.6 ^a	4.6 ^b	5.3	5.5 ^a	1.37
ATST-Effic.	3.8 ^b	3.9 ^b	4.7 ^a	4.9	5.5	5.5	5.3	5.6	5.8	84.90
ATST-Safety	.79	.75	.82	.89	.93	.95	.89	.93	.93	23.45
ATST-PA	1.6	1.8	1.8	1.8 ^b	2.1 ^a	1.9	1.9	1.9	1.9	16.67
Scan	6.5 ^b	6.5 ^b	7.6 ^a	7.3	7.8	7.8	7.5	7.7	7.9	15.27
Dial Reading	12.3 ^b	12.7	13.4 ^a	13.4	13.3	14.0	13.1 ^b	13.5	14.1 ^a	21.28
Analogies	8.3	7.9	9.4	8.9	8.7	9.8	9.3	9.1	9.9	17.4
EQ Composite	11.2 ^b	11.4 ^b	13.8 ^a	10.8 ^b	13.8 ^a	13.5 ^a	10.7 ^b	13.2 ^a	13.4 ^a	10.25

Note: Letters (a,b) following the means are indicative of post-hoc (between groups) test results of significantly different groups

ing index represent the standard deviation (in z-scores) change in score due to practice and coaching. Unweighted scores from all AT-SAT subtests were converted to z-scores. These unweighted scores were summed to calculate the AT-SAT unweighted composite. The AT-SAT weighted composite is the final score that is generated by the test software and used for selection decisions. As depicted in Figure 2, an individual-level practice index was created by subtracting the Time 1 score from the Time 2 score for participants from the control group. A group-level coaching index was calculated by subtracting the mean of the Time 1 scores for people who had not received coaching from the mean of the Time 2 scores of people who had been coached.

The mean of the Practice index for the weighted composite are each AT-SAT subtest is presented in Table 6. These results show that the AT-SAT composite scores increased by less than 1 standard deviation (SD) due to practice effects. This increase in score from Time 1 to Time 2 testing is not statistically significant. The AT Efficiency score is the AT-SAT sub-test that is most highly influenced by practice effects, increasing an average of .81 standard deviation from Time 1 to Time 2. Although the extent of the change varied as depicted in Table 6, these indices demonstrate the increase in performance that may be expected for each sub-test due to practice effects. In general, these results show that the computer simulation tests contained in the battery were more susceptible to practice effects.

The Coaching index for each unweighted AT-SAT test is presented in Table 7. These results show that the AT-SAT weighted composite score increased an average of .72 standard deviations due to coaching. Whereas performance on the computer simulation tests seems to be more susceptible to practice effects, coaching seemed to have a larger impact on non-cognitive test performance. As shown in Table 7, scores on the Experience Questionnaire (EQ) scales increased by as much as 1 standard deviation due to coaching. The greatest change in score was on the EQ consistency of work behavior scale, which increased by 1.20 standard deviations. The EQ interpersonal tolerance scale increased by 1.05 standard deviations, followed by EQ working cooperatively (coaching index=1.02), EQ decisiveness (coaching index=0.96), EQ Composure (coaching index=0.84), and EQ Self-confidence (coaching index=0.62).

Hiring Decisions

As stated above, another objective of the current study was to investigate the potential impact of practice and coaching effects on personnel hiring decisions. In essence, what are the practical implications of the practice and coaching effects that were found? Could the increases in performance alter hiring decisions and give an advantage to someone who received coaching? These issues were investigated using scores from both the coaching after testing and control groups once all data had been collected. This comparison focused on coaching rather than practice

Figure 2. Calculation of Practice and Coaching Index

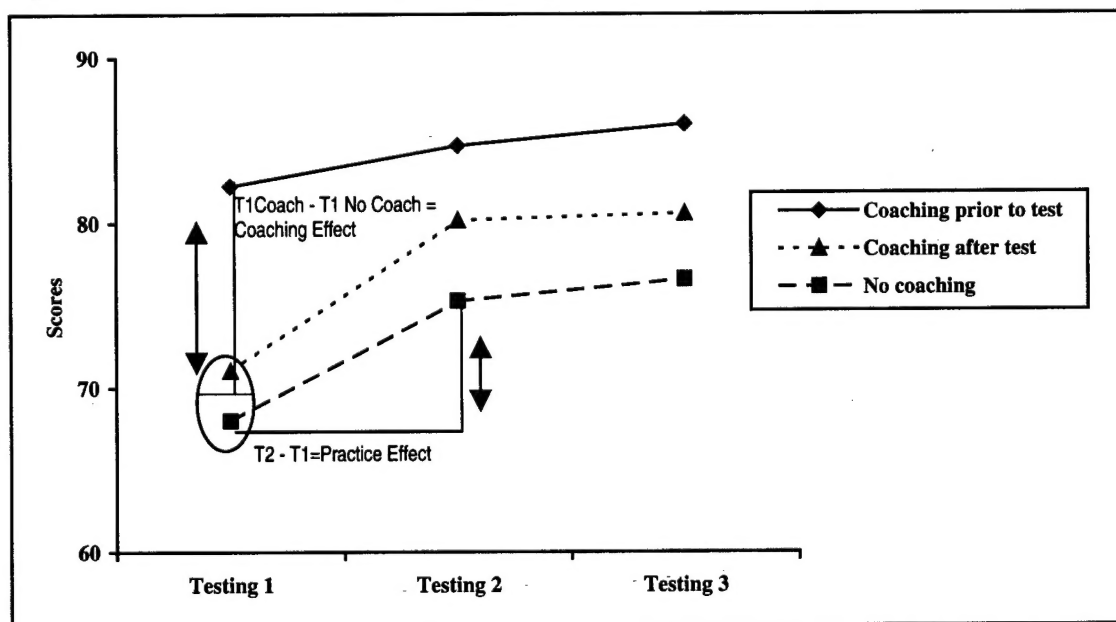


Table 6. AT-SAT Practice Index

AT-SAT (Unweighted Composite)	0.4336
AT-SAT (Weighted Composite)	0.4423
AT Efficiency	0.8103
LF- Situational Awareness	0.5430
Dial Reading	0.5030
AT Safety	0.4921
AT Procedural Accuracy	0.4064
Scan	0.3540
Analogies	0.2082
Angles	0.2002
Applied Math	0.1819
EQ Composure	-0.0085
EQ Self-confidence	-0.0096
EQ Interpersonal Tolerance	-0.0774
EQ Working Cooperatively	-0.0995
LF- Planning and Thinking Ahead	-0.1114
EQ Consistency of Work Beh.	-0.1591
EQ Decisiveness	-0.1638

Table 7. AT-SAT Coaching Index

AT-SAT (Unweighted Composite)	1.061
AT-SAT (Weighted Composite)	0.720
EQ Consistency of Work Beh.	1.2010
EQ Interpersonal Tolerance	1.0470
EQ Working Cooperatively	1.0185
EQ Decisiveness	0.9553
EQ Composure	0.8372
EQ Self-Confidence	0.6164
AT Efficiency	0.5696
LF- Situational Awareness	0.5370
Angles	0.4819
Scan	0.4763
LF- Planning and Thinking Ahead	0.4615
Applied Math	0.3340
Analogies	0.3157
Dial Reading	0.3059
AT Safety	0.2373
AT Procedural Accuracy	0.1728

because a practice effect existed for all participants once they take the test for the second time. However, due to the design of the study, not everyone had a coaching effect, enabling a direct comparison between participants who have been coached and those who have not. In essence, the ranking example contains both coaching and practice, yet the only variable that actually differs for the participants is whether or not they had been coached. In the real world, it is unlikely that all job candidates receive some type of coaching, so an understanding of the differential advantage gained by those who are coached is of practical importance. The procedures used to make this comparison are as follows:

- 1) Time 1 AT-SAT scores were rank-ordered for all participants who had not received coaching;
- 2) Time 2 AT-SAT scores, where one of the groups had received coaching just prior to taking the test, were then rank-ordered and compared with Time 1 rankings.

The results of this comparison are shown in Table 8. Since AT-SAT scores will be used by the FAA to rank order ATCS candidates for top-down selection, this process was replicated using the research participants. If a hiring decision were made after the first administration of AT-SAT, and prior to any coaching, three of the top five candidates would be from the control group (i.e. received no coaching throughout the study). Two of the

top five candidates belonged to the coached group (i.e. received coaching *after* Time 1 testing and immediately *before* Time 2 testing). As shown in Table 8, after Time 2 testing, several of the top ten candidates who had just received coaching improved their ranking and moved ahead of people who had received no coaching. Whereas three of the top five candidates at Time 1 were from the control group, only one of these people remained ranked in the top five after Time 2 testing. The remainder of the Top five candidates had all received coaching. The average increase in ranking for people who had been coached, based on weighted AT-SAT composite, was 6.23; the average for people who received no coaching was .64.

DISCUSSION

As previously stated, the goals of the current study were to: 1) determine if repeated test taking improves performance; 2) determine if coaching improves performance; 3) identify specific tests within the AT-SAT battery that are most susceptible to practice and coaching effects; and, 4) determine the extent to which practice and coaching effects potentially impact hiring decisions.

The results of this study identify which tests are most susceptible to both practice and coaching effects so that they can be monitored and targeted for alteration if needed. The results suggest that performance

Table 8. Change in Rank Order of Top Ten Candidates Following Coaching

Hiring Decision			
Treatment	Time 1 Subject	Time 2 Subject	Treatment
control	A	E	coached
coached	B	B	coached
control	C	H	coached
control	D	I	coached
coached	E	C	control
coached	F	A	control
coached	G	D	control
coached	H	F	coached
coached	I	G	coached
control	J	J	control

on the AT-SAT battery may indeed be influenced by both practice and coaching effects. The results of the ANOVA demonstrate that the composite AT-SAT score that is used for hiring decisions increases with repeated administrations, although the greatest increase occurs following coaching. The average increase in composite AT-SAT score due to coaching was greater than the average increase due to practice. The implications of this increase on hiring decisions are discussed later in this section.

With regard to the impact of practice, performance-based tests were affected more by practice than were tests that required knowledge or abilities not measured by computer simulations. However, this was not always the case. Performance on one non-simulation test, dial reading, improved by 1/2 a standard deviation due to practice. This may have occurred because the dial reading test is a relatively easy test to learn and any strategies learned during the first test session could be easily applied to subsequent sessions. Not all simulation scores improved with practice; performance on one of the Letter Factory scores (Planning and Thinking Ahead) actually decreased. Based on reports from study participants, the lower LF score may have been due to increased vigilance to the situational awareness aspect of the test, which hindered performance on the Planning and Thinking Ahead dimension.

With regard to the impact on the non-cognitive test (the EQ), it was most susceptible to coaching effects. The large increase in score on the EQ scales following

coaching suggests that the participants were able to easily learn how to fake well on this measure of personality in the workplace. Coaching also helped the study participants to improve some of their computer simulation test scores. Although all of the computer simulation scores improved after coaching, the largest increases on the cognitive tests occurred for both Letter Factory scores, one of the ATST scores, the Scan test, and the Angles test. One reason for the improvement for these particular tests may be that specific strategies taught for these tests may have been easier to remember and more effectively applied. The implication is that these particular tests are both more easily and more effectively coached.

Implications

In terms of implications for personnel selection decisions, seven of the people ranked among the top ten candidates received coaching. Of those people ranked in the top 5, only one had not received coaching. Further review of rankings revealed that the average increase in ranking for people who had been coached was much more dramatic than it was for those who did not receive coaching. The trend described above continues beyond the top third of the candidates. The focus in this report is on the top ten because it illustrates the impact on top down selection, particularly if you are only taking the top 3 to 5 candidates. The impact of coaching effects on personnel decisions will decrease as more people from the candidate pool are hired. However, as demonstrated by

these results, coaching may have a large impact on the selection decisions made by an organization that selects candidates using a top down approach. These results demonstrate that people who received coaching had an advantage over the uncoached people. Candidates who would have not been selected after the first test administration were chosen after the second, primarily because they had received coaching on how to take the test. Since the coaching was intended to improve scores without improving the cognitive abilities required for the job, it is unlikely that these increases in AT-SAT scores translates into improved performance on the job. The FAA does not currently use a strict top down approach, rather it uses a category grouping method. This method divides applicants who pass AT-SAT (that is, applicants who receive an AT-SAT score of 70 or better) into "qualified" (AT-SAT score of 70 – 84.9) and "well qualified" (AT-SAT score of 85 or higher) categories. Nevertheless, it is conceivable that coaching could move an individual from a failing status into a passing status and even from the qualified category into the well-qualified category, as well as improving their standings within these categories, again without improving their ability to perform on the job.

Limitations

There are a few important limitations of this study that must be considered when reviewing the results. Once AT-SAT becomes operational, candidates will be unable to retake that test for one year after their initial testing session. Due to limitations of time and concerns about attrition, the current study could not wait for one year between test administrations to assess practice effects. Consequently, the 4-week time interval between test administrations does not reflect the actual practice or policy that will be in place operationally. As such, the practice effects described in this study likely reflect the "worst case scenario" and overestimate the impact of practice after a 1-year time interval. An additional limitation pertains to the coaching used for this study. As described above, the coaching session was developed by ATCSs and CAMI researchers who had substantial knowledge of the AT-SAT battery. It is unlikely that any potential coaching sessions developed by third party vendors will contain the degree of accurate detail included in the coaching used for the current study. Consequently, the coaching effects described in this study are also likely to reflect the "worst case scenario" and overestimate the impact of coaching on test performance.

REFERENCES

- Alliger, G.M., Lilienfeld, S.O., & Mitchell, K.E. (1996). The susceptibility of overt and covert integrity tests to coaching and faking. *Psychological Science*, 7, 32-39.
- Furnham, A. (1997). *The Psychology of Behaviour at Work*. Hove East Sussex, UK: Psychology Press.
- Gilliland, K., & Schlegel, R. (1992). *Evaluation of extended practice effects on the Air Traffic Scenarios Test*. Unpublished manuscript.
- Houston, J.S., & Schneider, R.J. (1997). Analysis of Experience Questionnaire (EQ) Beta Test Data. Unpublished manuscript.
- Manning, C.A., & Heil, M.C. (2001). *The Relationship of FAA Archival Data to AT-SAT Predictor and Criterion Measures*. In Ramos, R., Heil, M.C., & Manning, C.A. (Eds.) (2001). *Documentation of Validity for the AT-SAT Computerized Test Battery*. (DOT/FAA/AM-01/6). Washington, DC: Federal Aviation Administration Office of Aerospace Medicine¹.
- Mauer, T., Solamon, J., & Troxtel, D. (1998). Relationship of coaching with performance in situational employment interviews, *Journal of Applied Psychology*, 83, 128-36.
- Nickels, B.J., Bobko, P., Blair, M.D., Sands, W.A., & Tartak, E.L. (1995). *Separation and Control Hiring Assessment (SACHA); Final job analysis report*. Washington, DC: Federal Aviation Administration Office of Personnel.
- Powers, D.E. (1993). Coaching for the SAT: A summary of the summaries and an update. *Educational Measurement: Issues and Practice*, 12 (2), 24-30.
- Sackett, P.R., Burris, L.R., & Ryan, A.M. (1989). Coaching and practice effects in personnel selection. In C. L Cooper, & I. T. Robertson, (Eds.) *International Review of Industrial and Organizational Psychology*. Chichester, NY: Wiley.

¹All Office of Aerospace Medicine technical reports are available in full-text from the Civil Aerospace Medical Institute's publications Web site: <http://www.cami.jccbi.gov/aam-400A/index.html>

APPENDIX A

Applied Math. This test contains 30 multiple-choice questions. The test presents five practice questions before the test begins. Questions such as the following are contained on the test: *A plane has flown for 3 hours with a ground speed of 210 knots. How far did the plane travel?* These questions require the participant to be able to factor in such things as time and distance in order to identify the correct answer from among the four answer choices. (Total Time: 30 minutes.)

Angles. The Angles test measures the participant's ability to recognize angles. This test contains 30 multiple-choice questions. There are two types of questions on the test. The first presents a picture of an angle and the participant chooses the correct size of the angle (in degrees) from among four response options. The second presents a measure in degrees and the participant chooses the angle (among four response options) that represents that measure. (Total Time: 10 minutes.)

Letter Factory Test (LF). This test simulates a factory assembly line that manufactures only four letters of the alphabet (A, B, C, and D) in one of three colors. The test has 18 sections and requires that participants use a mouse to perform multiple and often concurrent tasks. Each test section begins with letters appearing at the tops of the conveyor belts moving down toward the loading area. Based on those letters, participants immediately begin selecting and moving boxes to the loading area to provide just the right number and color of boxes to correctly place all letters. Other tasks performed during the simulated factory settings include: (1) picking up letters of various colors, (2) ordering new boxes when supplies become low, and (3) calling Quality Control when defective letters appear. Each section lasts between 30 seconds and 2 1/2 minutes. (Total Time: 91 minutes.)

Air Traffic Scenarios Test (ATST). This is a low-fidelity simulation of an air traffic control (ATC) radar screen that is updated every seven seconds. The goal is to maintain separation and control of a varying number of simulated aircraft (represented as data blocks) within the designated airspace as efficiently as possible. Aircraft in flight can pass through the airspace or land at one of two airports within the airspace. Each aircraft's data block indicates its present heading, speed, and altitude. There are eight different headings representing 45 degree increments, three different speeds (slow, moderate, fast), and four different altitude levels (1=lowest and 4=highest). Separation and control are achieved by communicating and coordinating with each aircraft by using the computer mouse to click on the data block representing each aircraft and providing instructions such as changes to current heading, speed, or altitude. (Total Time: 95 minutes.)

Scan. In the Scan test, participants monitor a field that contains discrete objects (called data blocks) which are moving in different directions. Data blocks appear and move in the field at random, then disappear. During the test, the participant sees a blue field that fills the screen, with the exception of a 1-inch white bar at the bottom. In this field, up to 12 green data blocks may be present. Each data block contains two lines of letters and numbers separated by a horizontal line. The upper line is the identifier and begins with a letter followed by a 2-digit number. The lower line contains a 3-digit number. Participants are scored on the speed with which they notice and respond to the data blocks that have a number on the lower line outside a specified range. Throughout the test, this range is displayed at the bottom of the screen (e.g., 360-710). To "respond" to a data block, the participant types the 2-digit number from the upper line of the block (ignoring the letter that precedes it), then presses "enter." (Total Time: 18 minutes.)

Dial Reading Test. The Dial Reading test is designed to test the participant's ability to quickly identify and accurately read certain dials on an instrument panel. Participants are asked to choose from one of five response alternatives for each question about a given display. The test consists of 20 questions. Individual items are self-paced against the display of time left in the test as a whole. Participants are advised to skip difficult items and come back to them at the end of the test. Each panel consists of nine dials in two rows, a layout which remains constant throughout the test. Each of the nine dials contains unique flight information. (Total Time: 12 minutes.)

Analogies. The Analogies test measures the participant's ability to apply the correct rules to solve a given problem as well as their efficiency in using the available information to solve that problem. Analogies are based on words, pictures, or figures and appear in three "windows" on the same screen for a given item. Participants use a mouse to move freely between the three windows, view the different parts of the analogy, and select their answer. However, they can view only one window at a time. Window A presents the first part of the analogy that requires participants to infer the underlying rule. Window B contains that second part of the analogy that requires participants to apply the inferred rule. Finally, Window C provides participants the opportunity to confirm their choice by selecting their answer from the available response options. The test has 57 items: 30 word analogies and 27 visual (i.e., either pictorial or figural) analogies. (Total Time: 45 minutes.)

Experience Questionnaire (EQ). The operational version of the EQ contains 135 items allocated across nine scales. The nine scales are Composure, Consistency of Work Behavior, Working Cooperatively, Decisiveness, Self-Confidence, Interpersonal Tolerance, Execution, Task Closure, and Unlikely Virtues. EQ items are written as statements about the examinees' past experiences. Response options include: definitely true, somewhat true, neither true nor false, somewhat false, and definitely false. Internal consistencies range from .66 to .85 (Houston, & Schneider, 1997).